

## Cloning and Expression of the $\beta$ -N-Acetylglucosaminidase Gene from *Streptococcus pneumoniae*

GENERATION OF TRUNCATED ENZYMES WITH MODIFIED AGLYCON SPECIFICITY\*

(Received for publication, November 18, 1994, and in revised form, February 8, 1995)

Valerie A. Clarke†, Nick Platt§, and Terry D. Butters†¶

From the †Glycobiology Institute, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, and the  
§Sir William Dunn School of Pathology, University of Oxford, OX1 3QX, United Kingdom

The gene encoding a  $\beta$ -N-acetylglucosaminidase from *Streptococcus pneumoniae* has been obtained by screening an expression library for  $\beta$ -N-acetylglucosaminidase activity. Clones of different nucleotide sizes each having arylglycoside activity were obtained, and DNA sequencing revealed a gene of 3933 base pairs possessing typical bacterial transcription initiation and termination sequences and terminating in an ochre stop codon. Computer analysis of the translated protein of 1311 amino acids (144,210 Da) identified a tandem repeat within which lies a sequence homologous with six other hexosaminidase gene products from a wide variety of species ranging from bacteria to humans. Also found were an amino-terminal putative secretion signal peptide and a carboxyl-terminal cell sorting/anchorage motif typically found in over 20 other Gram-positive surface proteins. The expression of an almost complete DNA clone in *Escherichia coli* produced a functional and authentic  $\beta$ -N-acetylglucosaminidase with aglycon specificity identical to the wild-type enzyme. However, enzymes produced from truncated DNA clones show more restricted aglycon specificity and are unable to hydrolyze terminal  $\beta$ 1-2GlcNAc residues from N-glycans containing a bisecting N-acetylglucosamine. The availability of these clones allows structural analyses to be made of catalytic and oligosaccharide recognition protein domains that enhance functional activity.

$\beta$ -N-Acetylglucosaminidase, found in both the culture medium and in association with undisrupted cells, is one of six extracellular *Streptococcus pneumoniae* glycosidases purified to date (1-3). This repertoire of enzymes, in addition to the  $\beta$ -N-acetylglucosaminidase, includes a  $\beta$ -galactosidase, three endoglycosidases, and a neuraminidase that are thought to aid the organism in the breakdown of oligosaccharides in its surrounding environment for use as a carbon source. In particular, the neuraminidase has been implicated in the process of pathogenesis. By cleaving the terminal sialic acids on cell-surface glycolipids, it is thought that the neuraminidase can expose the carbohydrate ligand by which the *S. pneumoniae* attaches to the host (4, 5). *S. pneumoniae* also synthesizes and secretes a

hyaluronidase that is capable of degrading a component of the extracellular matrix and may aid the organism to invade underlying host tissue (6). Since GlcNAc $\beta$ 1-linked residues are common components of several cell-surface molecules in host tissues, it is possible that  $\beta$ -N-acetylglucosaminidase also has a pathogenic role.

All of these glycosidases are important reagents for oligosaccharide analysis.  $\beta$ -N-Acetylglucosaminidase has been demonstrated to have broad specificity using GlcNAc-Gal, hydrolyzing both  $\beta$ 1-3 and  $\beta$ 1-6 linkages that are commonly found in mucins. By contrast, the hydrolysis of N-linked sugars at low enzyme concentrations is restricted to GlcNAc $\beta$ 1-2Man linkages except when the Man- $\alpha$ 1-6 arm is substituted with a GlcNAc at both C-2 and C-6 positions, or the Man  $\beta$ -linked to the chitobiose core is substituted at the C-4 position by a bisecting GlcNAc (7). This limited activity can be a useful tool in the sequencing of oligosaccharides, giving specific structural information about the cleaved bond and the surrounding sugar residues. It also raises the issue of defining the intrinsic properties of the enzyme that govern this phenomenon. The aim of this study was to use an expression cloning strategy to obtain the gene for  $\beta$ -N-acetylglucosaminidase and to investigate the interaction between the enzyme and its N-linked carbohydrate substrates at the molecular level. By the expression of truncated portions of the  $\beta$ -N-acetylglucosaminidase gene, we have been able to examine defined regions of the protein that determine glycosidase specificity.

### EXPERIMENTAL PROCEDURES

**Materials**—Oligonucleotides were synthesized by British Biotechnology (Oxford, United Kingdom). Restriction and DNA-modifying enzymes were purchased from Boehringer Mannheim and New England Biolabs. Radionucleotides were obtained from Amersham Corp. and ICN-Flow (High Wycombe, Bucks, UK). Prime-It random primer DNA labeling kit was purchased from Stratagene.

**Bacterial Strains**—XL1-Blue *Escherichia coli* strain (recA- (recA1, lac-, endA1, gyrA96, thi, hsdR17, supE44, relA1, (F' proAB, lacI<sup>q</sup>, lacZ $\Delta$ M15, Trn10))) was purchased from Stratagene. NM554 *E. coli* strain (recA13, araD139,  $\Delta$ (ara-leu)7696,  $\Delta$ (lac)17A, galU, galK, hsdR, rpsL(strr) mcra, mcrB) was obtained from Stratagene.

**Purification of  $\beta$ -N-Acetylglucosaminidase**—A homogeneous preparation of the cellular form of  $\beta$ -N-acetylglucosaminidase was obtained from *S. pneumoniae* (ATTC 12213) using previously published methods (3). A limited trypsin digest of  $\beta$ -N-acetylglucosaminidase followed by amino-terminal amino acid sequencing was used to obtain two tryptic peptide sequences, a major sequence, EGADIPGIIGMVVA, and a minor sequence, LQPMFND. Samples for SDS-polyacrylamide gel electrophoresis were prepared by methods previously described (3).

**Construction and Screening of Genomic Expression Library**—Genomic DNA was prepared from frozen *S. pneumoniae* cells (1 ml of packed cells) following a method used for tissue samples (8). Extracted genomic DNA was partially digested with *AluI* and separated by aga-

\* The Glycobiology Institute is supported by Monsanto/Searle Company. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) L36923.

¶ To whom correspondence should be addressed. Tel.: 44-865-275725; Fax: 44-865-275216.

rose gel electrophoresis. Fragments ranging between 1 and 7 kb<sup>1</sup> were gently extracted from the agarose by adsorption to glass beads using the GeneClean II Kit (Bio101, Inc., Stratech Scientific, Luton, UK) to avoid shearing. The ends of the size-selected pieces were then blunt-ended using Klenow and T4 DNA polymerase (8). *Eco*RI linkers were ligated to blunt-ended genomic DNA fragments with T4 DNA ligase, and these inserts were ligated into *Eco*RI-digested  $\lambda$ -ZapII vector arms (Stratagene) and packaged with Gigapack® II packaging extracts (Stratagene) to generate a  $\lambda$  bacteriophage expression library. The number of primary recombinants was determined by plating the library on the XL1-Blue *E. coli* host strain. The plated expression library was screened for  $\beta$ -N-acetylglucosaminidase activity by incorporation of 50  $\mu$ M 4-methylumbelliferyl N-acetylglucosaminide in the top agarose and visualization of hydrolysis by UV light. Positive clones were plaque-purified by successive plating, and pure plaques were subcloned into plasmid Bluescript SK- by the *in vivo* excision protocol that accompanies the vector (Stratagene).

**Rescreening of Library for Overlapping Clones**—The library was plated and screened by plaque hybridization (Colony/Plaque Screen nylon discs, DuPont NEN) using DNA probes that were ( $\alpha$ -<sup>32</sup>P)dATP-labeled using the Prime-It random primer labeling kit (Stratagene) and purified over Select-D<sup>®</sup>, G-25 spin columns (5 Prime  $\rightarrow$  3 Prime, Inc., CP Labs, Bishop's Stortford, Herts, UK). DNA probes were prepared from the 5' end of pBSstrH7 by restriction digesting with *Bgl*III and *Eco*RI and from the 3' end of pBSstrH17 by restriction digesting with *Nco*I and *Eco*RI to generate probes of 570 and 730 bp respectively.

**Plasmid DNA Purification**—Clones were amplified and purified using the maxi-prep pZ523<sup>®</sup> spin column plasmid purification kit (5 Prime  $\rightarrow$  3 Prime, Inc.). DNA was also purified using Magic Mini Preps (Promega).

**Southern Blot Analysis of Clones**—Southern transfer of DNA to nylon membranes (Hybond, Amersham Corp.) was made according to published methods (8, 9). Oligonucleotide probes designed from the amino acid sequence of the major tryptic peptide from the purified protein, GAA GG(T/A) GC(T/A) GAT AT(T/C) CC(AT) AT(T/C) AT(T/C) GG(T/A) GG(T/A) ATG GT, were labeled with [ $\gamma$ -<sup>32</sup>P]dATP using polynucleotide kinase (8). After hybridization at 56 °C overnight, filters were washed and subjected to radioautography.

**DNA Sequencing**—Nucleotide sequences were determined using the dideoxy chain termination method (10) using the Sequenase DNA sequencing kit (U. S. Biochemical Corp. Cambridge BioScience, Cambridge, UK). Both strands of DNA were sequenced in their entirety by a combination of specific oligonucleotide primers and the transposon-facilitated DNA sequencing TN1000 nested set kit (Gold BioTechnology, Inc., St. Louis, MO). Sequence analysis was accomplished using MacVector and Assembly-Lign software (IBI, Cambridge, UK) and on-line NCBI data bases.

**Northern Blot Analysis**—Total RNA was prepared from *S. pneumoniae* using guanidinium thiocyanate (11). RNA (10  $\mu$ g) was separated on a denaturing formaldehyde agarose gel with RNA molecular sizing fragments ranging from 0.24 to 9.5 kb and Northern blotted. The nylon filter was probed with a 570-bp <sup>32</sup>P-labeled *Bgl*III-*Eco*RI fragment used to rescreen the genomic library.

**Subcloning into the pGX Expression Vector**—The vector pGEX-3X (Pharmacia, Milton Keynes, UK) was digested with *Eco*RI and dephosphorylated. The inserts from clones pBSstrH7, -8, and -17 were prepared by digestion with *Eco*RI and gel purification. The insert and vector were ligated overnight and transformed into NM554 *E. coli* by electroporation. Colonies were ampicillin-selected on LB/amp plates, and 20 were randomly chosen for analysis. Mini-prepped DNA was digested with *Eco*RI and separated on an agarose gel for verification of vector and insert sizes.

**pGEX Expression and Affinity Purification**—The expression and purification were carried out according to published methods (12). 50-ml cultures of *E. coli* transformed with pGEX clones were induced by the addition of isopropyl-1-thio- $\beta$ -D-galactopyranoside (1 mM), and the cells were sonicated. Affinity beads were mixed with bacterial cell lysates for 2 h at 4 °C and washed 3 times with buffer, and the fusion protein was eluted from the affinity beads by the addition of fresh 5 mM reduced glutathione. Where appropriate, factor Xa cleavage was conducted by incubating factor Xa (Denzyme, Aarhus, Denmark) at 1 mg/ml with beads coupled to fusion protein at room temperature overnight. Super-

natants were assayed for enzyme activity after centrifuging the beads.

**Substrate Specificity Assays**—Tritium-labeled biantennary, triantennary, tetraantennary, bisected-biantennary, and bisected-hybrid oligosaccharide alditols were obtained from Oxford GlycoSystems (Abingdon, UK). (GlcNAc $\beta$ 1,4GlcNAc)<sub>2</sub> was obtained from a partial chitin hydrolysis, and a degalactosylated biantennary native oligosaccharide was purified from human asialotransferrin (13).

Native oligosaccharide substrate was incubated at different concentrations (0.2–1.0 mM) with 2.5 milliunits/ml of wild-type or recombinant  $\beta$ -N-acetylglucosaminidase in 50 mM citric acid/sodium phosphate buffer, pH 5.0, containing 1 mg/ml bovine serum albumin at 37 °C for 1 h. The reactants were desalted, and hydrolysis was monitored by Dionex HPAEC (Dionex BioLC system) using a CarboPac PA-1 column eluted at 1 ml/min with 150 mM NaOH, 30 mM NaOAc, and the reaction products were detected using triple-pulsed amperometric detection with the following pulse potentials and durations:  $E_1 = 0.01$  V ( $t_1 = 120$  ms),  $E_2 = 0.6$  V ( $t_2 = 120$  ms), and  $E_3 = -0.93$  V ( $t_3 = 130$  ms). The extent of hydrolysis was calculated from empirically derived response factors for substrate and reaction products, and the data were plotted using a weighted nonlinear regression analysis (Multifit 2.0, Day Computing, Cambridge, UK). Radiolabeled oligosaccharide alditols were separated using an isocratic eluant of 200 mM NaOH and the fractions taken for radioactivity determination by scintillation counting. Bio-Gel P-4 chromatography (Oxford GlycoSystems) was also used to separate the reaction products after enzyme digestion, and the radioactivity in each fraction was determined as above.

## RESULTS

**Expression Cloning of *StrH***—An expression cloning strategy was adapted from published procedures (14) to obtain the gene encoding *S. pneumoniae*  $\beta$ -N-acetylglucosaminidase. A *S. pneumoniae* expression library of  $5 \times 10^6$  primary recombinants, generated in the vector  $\lambda$ -ZapII (Stratagene), was amplified in the *E. coli* host strain XL1-blue, and 300,000 clones were screened for  $\beta$ -N-acetylglucosaminidase activity using the substrate 4-methylumbelliferyl N-acetylglucosaminide (4-MU-GlcNAc). 32 positive clones were identified by fluorescent halos encircling each plaque when visualized under 366-nm ultraviolet light. Hydrolysis was confirmed to be specific for N-acetylglucosaminide by the inclusion of a control substrate 4-methylumbelliferyl xyloside in the screening protocol. No hydrolysis of this substrate was seen. From the group of positive clones, 20 were selected at random and subcloned into pBluescript plasmids via *in vivo* excision according to the manufacturer's procedures and analyzed by restriction digestion to determine the sizes of the genomic DNA inserts. Restriction-digested DNA of the four largest inserts ranging in size from 1584 to 2504 bp were Southern blotted and probed with a degenerate <sup>32</sup>P-labeled oligonucleotide designed from the amino acid sequence of the major tryptic peptide (Fig. 1). Hybridization of the radio-labeled probe to each clone selected by the enzyme activity of its translated product confirmed that they encoded the same protein that had been biochemically purified from the *S. pneumoniae* cells (Fig. 1). These clones (Fig. 2a) were nucleotide sequenced and were found to compose a 2.7-kb continuous open reading frame with no start or stop codons, indicating that further screening of the genomic library was required to locate the missing gene sequences. <sup>32</sup>P-Labeled DNA probes from the 5' and 3' ends of this 2.7-kb partial gene were used to identify two clones that contained the 5' and 3' ends of the *strH* gene (Fig. 2b). Complete sequence analysis of these three contiguous clones revealed a single, continuous open reading frame of 3933 bp designated *strH* (*str*, *streptococcus*; *H*, hexosaminidase) (Fig. 3). The ATG (methionine) start codon, designated +1, was identified by the positions of three consensus sequences: the AGGAGG Shine-Dalgarno ribosome binding sequence located 7 bases upstream from the ATG translation initiation codon, and the two putative promoter sequences TTGACT resembling a "[minus]35" transcription initiation sequence beginning at -87 and TATAAT "[minus]10" transcription initiation sequence beginning at -43. The ochre stop codon located at base 3934 was

<sup>1</sup> The abbreviations used are: kb, kilobase pair(s); StrH, *S. pneumoniae* hexosaminidase; 4-MU-GlcNAc, 4-methylumbelliferyl N-acetylglucosaminide; bp, base pair(s); HPAEC, high performance anion-exchange chromatography.

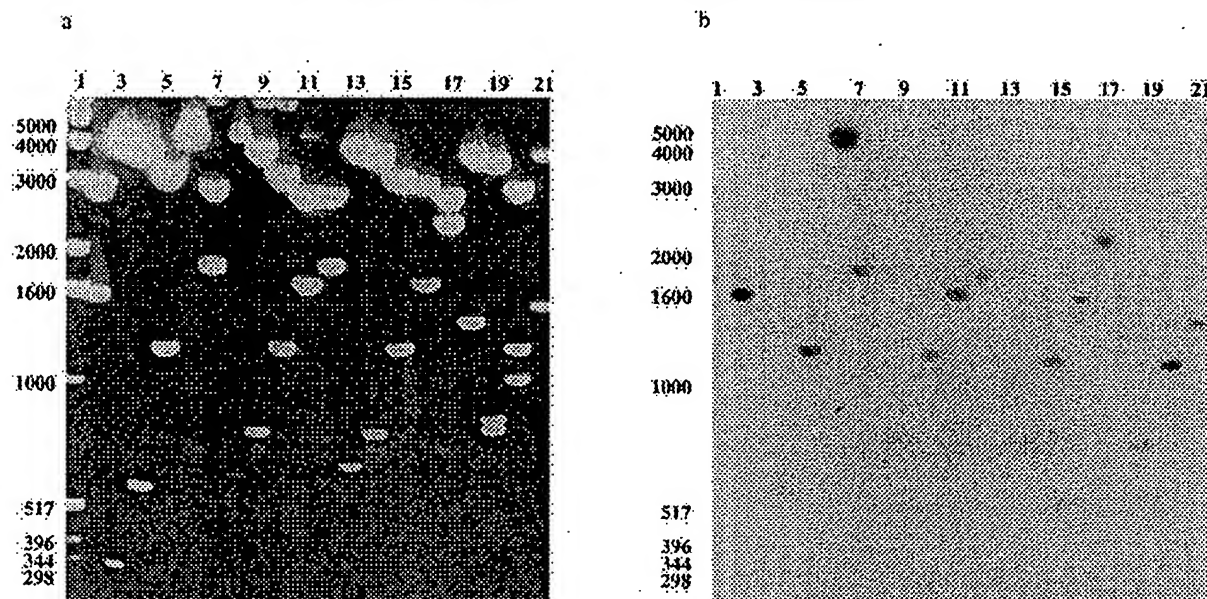
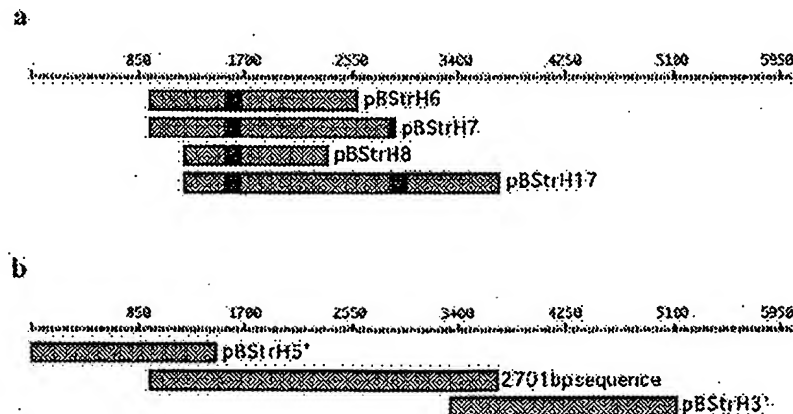


FIG. 1. Southern blot of *strH* clones. Clones 6, 7, 9, and 17 were restriction-digested with the following panel of enzymes: *Eco*RI, *Eco*RV, *Hind*III, *Cla*I, *Sac*I. DNA (1  $\mu$ g) was applied to a 0.7% Tris borate/EDTA buffer agarose gel in the order of enzymes listed above. Panel a shows the ethidium bromide-stained gel. Lane 1, BRL 1-kb DNA ladder; lanes 2-6, pBStrH6; lanes 7-11, pBStrH7; lanes 12-16, pBStrH9; and lanes 17-21, pBStrH17. In panel b the gel was transferred to a nylon membrane by Southern blotting and hybridized with a  $^{32}$ P-labeled degenerate oligonucleotide probe designed from the amino acid sequence of the major tryptic peptide. Filters were then washed twice in  $6 \times$  SSC at  $37^\circ\text{C}$ , once in  $4 \times$  SSC at  $57$ ,  $42$ ,  $47$ , and  $52^\circ\text{C}$ , and once in  $2 \times$  SSC at  $52^\circ\text{C}$  with cautious monitoring before subjecting to autoradiography.

FIG. 2. Alignment of overlapping clones. a, the clones, selected by screening for  $\beta$ -N-acetylglucosaminidase activity, were sequenced and aligned using MacVector and Assembly-Ign software. The relative position of the 96-bp tandem repeating regions encoding 30-amino acid consensus sequences to other hexosaminidases (Table 1) is shaded. b, the relation is illustrated between the contiguous 2701-bp partial gene composed of pBStrH6, -7, -9, and -17 obtained by activity screening of the genomic library and the pBStrH5' and -3' clones identified by screening the library with DNA probes derived from the ends of the contiguous sequence.



followed by a hairpin-loop sequence resembling a rho-independent transcription termination sequence.

**Protein Sequence of  $\beta$ -N-Acetylglucosaminidase.**—The *strH* DNA was computer translated to give a protein of 1311 amino acids with a predicted molecular mass of 144,210 Da. Both tryptic peptide sequences obtained from a purified preparation of  $\beta$ -N-acetylglucosaminidase were found (Fig. 3), and a comparison with all other published sequences in computer data bases did not reveal significant homology, confirming that the *strH* gene was unique. However, two amino acid sequence motifs common to other Gram-positive surface proteins were discovered (15). The first was a putative secretion signal peptide at the amino terminus revealed by Kyte-Doolittle hydrophobicity analysis (Fig. 4a). This sequence motif is characterized by a positively charged amino terminus followed by a hydrophobic core and a string of polar residues (16). The proposed cleavage site of this signal peptide is between the two alanine residues at sites 23 and 24 in *strH* between the hydrophobic core region and the polar residues (17). Within this

region is a KQQRFSIRKXXXGAASLIQ consensus sequence that appears to be highly conserved in many of the cell surface proteins of *Streptococcus* and *Staphylococcus* (Table 1). The carboxyl terminus contained a sequence homologous to a membrane spanning/cell wall anchorage sorting signal found in more than 20 Gram-positive surface proteins to date (18) (Table 1), identified by the consensus amino acid sequence LPXTGX, where LP and G are the conserved amino acids. In the translated sequence, we obtained LPETGTF, followed by a hydrophobic stretch of amino acids proposed to be the membrane spanning region, and a short charged tail (Fig. 4b). The LPETGTF sequence is directly preceded by two lysine residues that may be of importance in anchoring the enzyme to the cell wall. It has been suggested (18) that the two lysines preceding the LPXTGX motif of protein A may be covalently attached via a transpeptide linkage to the peptidoglycan of the cell wall of *Staphylococcus aureus*. The authors based this proposal on the finding that the murein of the peptidoglycan in *E. coli* is covalently attached to lipoprotein via lysine and arginine resi-

-773 CGCTCTAGAACTAGTGGATCCCGGGCTGCAGGAAATGGCAGCACTGGTTGTTACTTGG  
 -713 AAAAAATGGGTCTACGGCGGATCAATCTACTACAGATTGAACGTATCCGCGAATCAA  
 -653 ACGGTGTCGTTTATATCTACCATACCTTACATCCAGAAACATATATCAACGCAATTT  
 -593 ATCCAAACCTTGAACTATAGTTCTATCTATAATCGTTTCAAGTTAGACTACCAATTC  
 -533 ACATGAATGATGAACACTTTGAAGAAATCAACGAATAGTCTTTCCAACTCCAGAACATG  
 -473 CGGCTTCAGTCTCGGAGTCGATGAACAATTCCTTACTGCTTACAACTAAATACCCA  
 -413 AACTCGAATCAACTGGCCAAATTTTGAATACAGTGAACCTTACAAACGATCTGATTACT  
 -353 ACAAGATTAAATTCATCTCATGTGACCGTGATCACTAAGAAGAAAGCCTGAGCCTAATCG  
 -293 CTCGGGCTTTTCTATGCTTATTATACACACCAAAAGGTAGCGCTTACCTATTITTTGG  
 -233 AAGATTCTATAATTTGTGGAATATTTATAGAAATAGAACCAAAATTTATCTAGATATA  
 -173 TATACGTATTTTTCGACCAATCTGAATCTTAAAGCTTAAATTTTATATTTTATATAC  
 -113 TTATTATAGTTTGTGTCATCTCTCTTGACTCTCGTTAATATAAGAAATAAATAAGG  
 -53 CGTTGATTATATAATCGCTATCAATATAACAAATCAATCAGGAGTTTTCGAATGAAC

M K H

8 ATGAAAAACAACAGCGTTTCTTATTCGTAATACGCTGTAGGAGCAGCTTCTGTTCTAA  
 E K Q Q R P S I R K Y A V G A A S V L I  
 68 TTGGATTGCGCTTCCAAGCAGAGACTGTTCAGCGGATCGAGTTTACTACTACAGAAA  
 G F A F Q A Q T V A A D G V T T T T E N  
 128 ACCAACCGACCATACAGTTTCTGATTCCTCAATCATCCGAAATCGGATCGAGG  
 Q P T I H T V S D S P Q S S E N R T E E  
 188 AAACACCTAAGCAGAGCTTCAACCAAGCTCCAAAACTGTAGAAACAGAACTCCAG  
 T P K A E L Q P E A P K T V E T P A  
 248 CTACTGATAAGTGTAGTCTTCCAAAAACAGAAAGAAACCAAGAGGAGGTAGTT  
 T D K V A S L P K T E E V K P S S  
 308 CAACTCCTAGTGATAAAGCAGAGGTGGTAACCTCAACTTCTGCTGAAAAAGAACTGCTA  
 T P S D K A E V V T P T S A E K E T A N  
 368 ATAAAAAGGCAGAGAGCTAGCCCTAAAAAGGAAGAGCGAAAGAGGTGTATCTAAAG  
 K K A E E A S P K K E E A K E V D S K E  
 428 AGTCAATACAGACAGACTGACAAGGATAAACAGCTAAAAAGATGAAGCGAAGCAG  
 S N T D K T D K D K P A K K D E A K A E  
 488 AGGCTGACAAACCGGAAACAGAGGCAGGAAAGCAAGCTGCTGCACTTAATGAAGAAC  
 A D K P E T E A G K E R A A T V N E K L  
 548 TAGCGAAAAAGAAATTTGTTCTATGATGCTGGACGTAATATTTCTCACCAGAACAC  
 A K K K I V S I D A G R K Y F S P E Q L  
 608 TCAAGGAATCATCGATAAAGCAACATTATGCTGCTACACTGATTTACACCTATTAGTCG  
 K E I I D K A K H Y G Y T D L H L V G  
 668 GAAATGATGGACTCCGTTTCTATGTTGGACGATATGAGCATACAGCTAACCGCAAGACCT  
 N D G L R F M L D D M S I T A N G K T Y  
 728 ATGCCAGTGACGATGTCAAACGCCCATTTGAAAAAGGTACAAATGATTATTACACGATC  
 A S D D V K R A I E K G T N D Y Y N D P  
 788 CAAACCGCAATCACTTAACAGAAAGTCAATGACAGATCTGATTAACTATGCCAAGATA  
 N G N H L T E S O M T D L I N Y A K D K  
 848 AAGGTATCGTCTCATTCGACAGTAATAGTCTGGACACATCGATCGGATCTCAATG  
 Q I G L I P T V N S P G H M D A I L N A  
 908 CCATGAAGAATTTGGAAATCCAAACCTTAACCTTATGCTATTTTGGGAAGAAATCAGGCC  
 M K E L G I O N P N F S Y F G C K K S A R  
 968 GTACTGCGATCTTGACAAAGCAAGCTGTCGCTTTTACAAAGCCCTTATCGACAAGT  
 T V D L D N E Q A V A F T K A L I D K Y  
 1028 ATGCTGCTTATTTGCGAAAAAGACTGAAATCTTCAACATCGGACTGTGATGAATATGCCA  
 A A Y F A K K T E I F N I G L D E Y A N  
 1088 ATGATCGCAGATGCTAAAGGTTGGAGTGTGCTTCAAGCTGATTAATATCTTCAACG  
 D A T D A K G W S V L Q A D K Y Y P N E  
 1148 AAGGCTACCTGTAAAGGCTATGAAAAATTTATGCTTACGCCAATGACCTCGCTCGTA  
 G Y P V K Q Y E K F I A Y A N D L A R I  
 1208 TTGTAAATCGCAGGCTCTCAACCAATGGCTTTTAAAGCGGTATCTACTACAAATAGCG  
 V K S H G L K P M A F N D G I Y Y N S D  
 1268 ACACAAGCTTTGGTGTGTTTACAAAGACATCATCGTTTCTATGCTGAGCTGTGCTGCG  
 T S P G S P D K D I I V S M W T G G W G  
 1328 GAGGCTACGATGCTGCTTCTTAACTACTAGCTGAAAAAGGTCAACCAATCGTTAATA  
 G Y D V A S S K L L A E K G H Q I L N T  
 1388 CCAATGATGCTTGGTACTAGCTTCTTGGACGAAAGCGTGATGGCAAGGCTGGTACAAATC  
 N D A W Y V L G R N A D G Q W Y N L  
 1448 TCGATCAGGGCTCAATGCTATTAACCAACCAATCACTTCTGTACCAAAACAGAAAG  
 D Q G L N G I K N T P I T S V P K T E G  
 1508 GAGCTGATATCCCAATCATCGGTGCTATGCTGCTGCTGGCTGACACTCCATCTGCAC  
 A D I P I I G G M V A A W A D T P S A R  
 1568 GTTATTCACCATCAGGCTCTTCAAACTCATGCTCATTTTGCATAATGCCAAGCTGAAT  
 Y S P S R L P K L M R H P A N A N A E Y  
 1628 ACTTCGAGCTGATTATGAATCTGCAGAGCAAGCACTTAACGAGGTACCAAGACCTGA  
 F A A D Y E S A E Q A L N E V P K D L N  
 1688 ACCGTTATCTGCAAGAGCTCAGCGCCGTAAAGAAAGCTGAAAAAGCTATTCGCTCTC  
 R Y T A E S V T A V K E A E S R L  
 1748 TCGATAGCAACCTTAGCCGTGCCCAACAGATACGATTGATCAAGCCATGCTTAACTTC  
 D S N L S R A Q Q D T I D Q A I A K L Q  
 1808 AAGAACTGTCAACAACTTGACCTCAGCCTGAAGCTCTAAAGAAAGAAAGCTAAAC  
 E T V N N L T L T P E A L K E E A K R  
 1868 GTGAGGTGAAAAACTTGCCAAAAACAGGTAACTCTCAATCGATGCTGGACGCAATACT  
 E V E K L A K N K V I S I D A G R K Y F  
 1928 TTACTCTGAACAGCTCAACCGCATCTAGACAAAGCCAGTCAAGCTCGGATATTCTGATG  
 T L N Q L K R I V D K A S E L G Y S D V  
 1988 TCCATCTCTTCTAGGAAATGACGGACTTCGCTTCTACTCGATGATGATGACCACTACTG  
 H L L L G N D G L R F L L D D M T I T A  
 2048 CCAACCGAAAAACCTATGCTAATGATGACGTTAAAAAGCTATTAAGGAACTAAAG  
 N G K T Y A S D D V K K A I I E G T K A

FIG. 3. strH DNA and amino acid sequence and flanking regions. *Eco*RI and *Alu*I sites flank both ends of the genomic sequence. Both tryptic peptide sequences are marked in **boldface**. Transcription initiation and termination sequences are double underlined. Shine-Dalgarno sequence is underlined. The start and stop codons are marked by a dotted underline, and the A of the ATG start codon is designated +1. The two 30-amino acid tandem repeating regions homologous with other hexosaminidases (Table II) are underlined. This sequence has been deposited in the GenBank data base (accession number L36923).

dues (19). Another striking feature found within the amino acid sequence of the  $\beta$ -N-acetylglucosaminidase is a tandem repeat of approximately 335 amino acids from site 180 to 522 and

again between sites 625 and 956. This repeat is apparent in both nucleotide sequence and amino acid sequence self-alignments, although greater homology is conserved when compar-

2108 CTTACTACGACGATCCAAACGGTACTGCACTAACACAGCCAGAACTAACAGAGCTAATTG  
 Y Y D D P N G T A L T Q A E V T E L I E  
 2168 AATACGCTAAATCTAAGGACATCGGTCTCATCCAGCTATTAAACAGTCCAGGTACATGG  
 Y A K S K D I G L I P A I N S P Q H M D  
 2228 ATGCTATGCTGGTGGCATGGAAAAATTAGGTATTAAAAATCCTCAAGCCCACTTTGATA  
 A M L V A M E K L G I K N P Q A H P D K  
 2288 AAGTTTCAAAAACTATGGACTTGA AAAACGAAGCGATGA ACTTTGTAAAAGCCC  
 V S K T T M D L K N E E A M N F V K A L  
 2348 TCATCGGTAAATACATGGACTTCTTTGAGGTAAAAACAAGATTTTCAACTTTGGTACTG  
 I G K Y M D F P A G K T K I F N P G T D  
 2408 ACGAATACGCCAAGATGCGACTAGTGGCCCAAGGCTGGTACTACCTCAAGTGGTATCAAC  
 E Y A N D A T S A Q G W Y Y L K W Y Q L  
 2468 TCTATGGCAAAATTTGCCGAATATGCCAACCCCTCCAGCTATGCCAAAAGAAAGAGGGC  
 Y G K F A E Y A N T L A A M A K E R G L  
 2528 TTCAACCAATGGCCCTCAACGATGGCTTCTACTATGAAGACAAGGAGATGTTCAAGTTTG  
 Q P M A F N D C P Y Y E D K D D V Q P D  
 2588 ACAAGATGTCTTGTATTCTTACTGGTCTAAAGGCTGGTGGGGATATAACCTCCGATCAC  
 K D V L I S Y W S K G W W G Y N L A S P  
 2648 CTCATACCTAGCAAGCAAGCCCTATAAATCTTGAATACCAACGGTACTGGTACTACA  
 Q Y L A S K G Y K P L N T N G D W Y Y I  
 2708 TTCTTGGTCAAAAACCAAGATGGTGGTGGTTTCCCTCAAGAAAGCTATTGAGAATACG  
 L G Q K P E D G G G F L K K A I E N T G  
 2768 GAAAAACACCATTCATCAACTAGCTTCTACCAAAATATCTGAAAGTAGATCTTCAACAG  
 K T P F N Q L A S T K Y P E V D L P T V  
 2828 TCGGAAGTATCTTCAATCTGGCCAGATAGACCAAGCCCTCAGTACAAGGAAGAGGAAA  
 G S M L S I W A D R P S A E Y K E E E I  
 2888 TCTTTGAAGTCTGACTGCTTTGACAGCCACAACAAAGACTACTTCCGTCTAATATATA  
 F E L M T A F A D H N K D Y F R A N Y N  
 2948 ATGCTCTCCGGAAGAAATTAGCTAAAATCTTACAACTTAGAAGGATATAGTAAAGAAA  
 A L R E E L A K I P T N L E G Y S K E S  
 3008 GTCTTGAGGCCCTTGACGCGCTAAAACAGCTCTAAATTACAACCTCAACCGTAATAAAC  
 L E A L D A A K T A L N Y N L N R N K Q  
 3068 AAGCTGAGCTTGACACACTTGTAGCCAACTAAAAGCCGCTTCTCAAGGCCCTCAAAACGAG  
 A E L D T L V A N L K A A L Q G L K P A  
 3128 CTGCAACTCATTCAGGAAGCCTAGATGAAAATGAAGTGGCTGCCAATGTTGAAACCGAGC  
 A T H S G S L D E N E V A A N V E T R P  
 3188 CAGAACTCATCAAGAACTGAAGAAATTCATTGAAAGTTATCAAGAAAGAAAATCTTA  
 E L I T R T E E I P F E V I K K E N P N  
 3248 ACCTCCAGCTGGTCAAGAAAATATTATCAGCAGGAGTCAAGGTGAACGAAGTCAAT  
 L P A G Q E N I I T A G V K G E R T H Y  
 3308 ACATCTCTGTACTCACTGAAAATGAAAAACAAGAAAGTCCCTTGATAGCCAGGTAA  
 I S V L T E N G K T T E T V L D S Q V T  
 3368 CCAAGAAAGTTATAAACCAAGTGGTTGAAGTTGGCGCTCTCTGTAACCTACAAGGTGATG  
 K E V I N Q V V E V G A P V T H K G D E  
 3428 AAAGTGGTCTTGACCAACTACTGAGGTAAAACCTAGACTGGATATCCAAAAAGAAAGAAA  
 S G L A P T T E V K P R L D I Q K E E I  
 3488 TTCCATTTACCACAGTGAAGTGAAGAAATCCACTCTTACTCAAGGAAAGAAACACAACTCA  
 P F T T V T R E N P L L L K G K T Q V I  
 3548 TTACTAAGGGCTCAATGAGCATCGTAGCAACTTCTACTCTGTGAGCACTTCTGCCGATG  
 T K G V N G H R S N F Y S V S T S A D G  
 3608 GTAAGGAAGTGAACAACTTGTAAATAGTGTGTCAGCACAGGAAGCCGTACTCAAAATAG  
 K E V K T L V N S V V A Q E A V T Q I V  
 3668 TCGAAGTCCGAACATATGGTAACATGATAGGCGATGAAAACGGACAAGCCGCTATTGCTG  
 E V G T M V T H V G D E N G Q A A I A E  
 3728 AAGAAAAACAAAAGTGAAGTCCCAAGCCAAACAGCTCCATCAACTGCTCTCTGCTGAGG  
 E K P K L E I P S Q P A P S T A P A E E  
 3788 AAAGCAAAGCTCTTCTCAAGATCCAGCTCTGTGGTAACAGAGAAAAAACTTCTGAAA  
 S K A L P Q D P A P V V T E K K L P E T  
 3848 CAGGAAGTCAAGATTCGAGGACTAGTACTCGCAGGACTCATGTCCACACTAGCAGCCT  
 G T H D S A G L V V A G L M S T L A A Y  
 3908 ATCGACTCACTAAAAGAAAAGACTAAGTCTTTTCGATAAAAAATAAACAGCGAGATT  
 G L T K R K E D \*  
 3968 GAAGCTCGCTGTTTATTTTAAATTAATACCTAGTCCAAGACGTTCAAAGATATCATCC  
 4028 ACTCGTTTGGTGTAAATAAAGTGGTGAAGATTTCACTCGATTCTTCTTGAGTGAGACGT  
 4088 GATGTTACTTCTGAATCTGCCTCAAGAACTGGTTTAAAGTCTACTTGGTGTCCCAAGAG  
 4148 TAGGCTGTTTTTGGTGGCAAGTCAATAGGCTTCGCTACGGGTCAATGCTTTTCAATC  
 4208 AATGTCACATAGCCCGTTGGCTAAAGATAAGACAAAAGTTCGAGTTTCATGTTTTCGGATC  
 4268 ATATTTTTCGGGAAGACTGCAAGTCTTGACGATATTTCCAAAACGGTTGAGCATGTAG  
 4328 TCAATCAAAATGCTGATCTGGTGTGATGATACGCTCACTCGTGGCCGAAATC

FIG. 3—continued

ing the amino acid sequences than is found between the nucleotide sequences in these repeat regions. Within each of these tandem repeat regions lies a sequence spanning 30 amino acids which, when compared to each other, are 67% identical and 93% similar when considering conservative amino acid substitutions. Interestingly, both of these 30 amino acid sequences have considerable homology to protein sequences found in six other hexosaminidases isolated from a wide variety of species including bacteria and humans (Table II).

**Northern Blot Analysis**—A total RNA preparation from *S. pneumoniae* was used to identify the transcript of  $\beta$ -N-acetylglucosaminidase. A major band hybridizing at 4.0 kb (results not shown) was consistent with the size of the 4.0-kb DNA

obtained from sequencing.

**Kinetic Analysis of Recombinant and Wild-type  $\beta$ -N-Acetylglucosaminidase**—Clones pBStrH8 and -17, as expected, showed arylglycoside hydrolytic activity, but to determine that these had the same activity as wild type  $\beta$ -N-acetylglucosaminidase using a native biantennary N-glycan, they were subcloned into the pGEX-3X expression vector (Pharmacia Biotech Inc.) for the high-level production of recombinant enzyme. The  $K_m$  and  $V_{max}$  values were determined for the two recombinant enzymes from clones pGXStrH17, the longest clone, and pGXStrH8, the most truncated clone. Wild-type enzyme demonstrated the highest affinity for the natural oligosaccharide with a  $K_m$  of 132  $\mu$ M compared with the  $K_m$  values of the

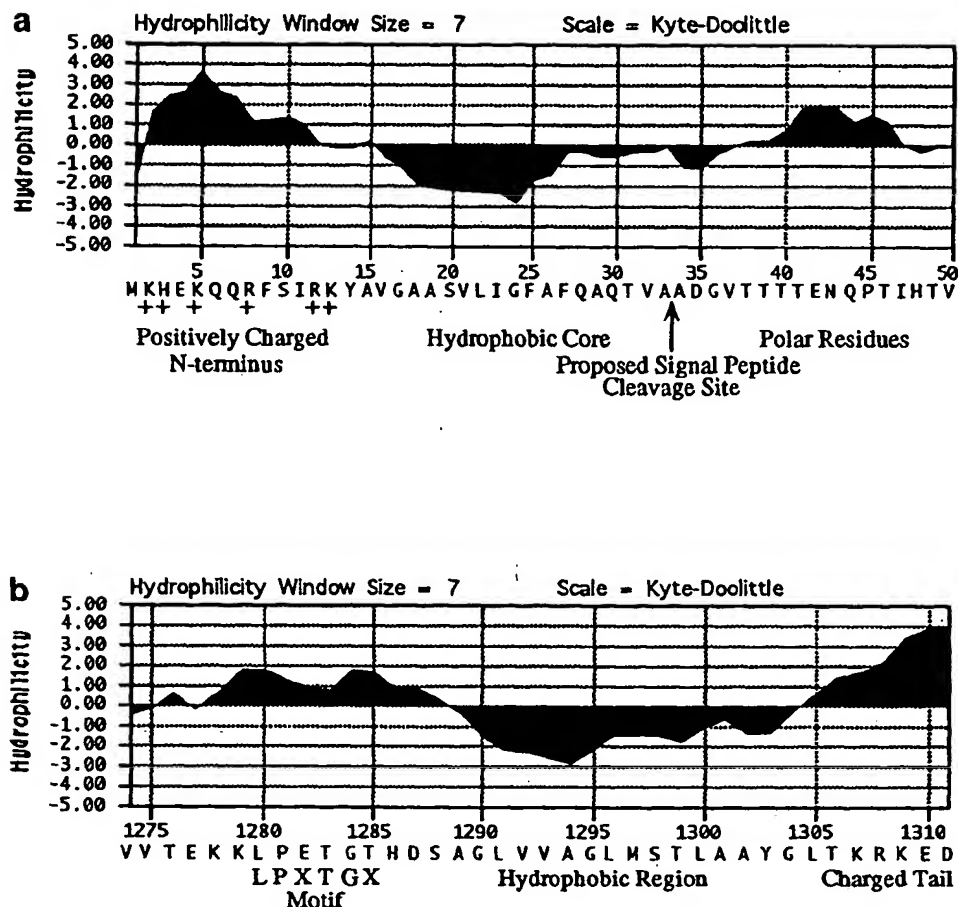
Cloning of the *Streptococcus*  $\beta$ -N-Acetylglucosaminidase Gene

FIG. 4. *a*, hydrophilicity plot of the amino terminus of *strH*. The *strH* translation product was analyzed by a Kyte-Doolittle hydrophilicity plot. The first 50 amino acids of the amino terminus were analyzed for features of a signal peptide motif found in other prokaryotes. The consensus motifs are labeled. *b*, hydrophilicity plot of the carboxyl terminus of *strH*. The last 35 amino acids of the *strH* carboxyl terminus were scanned for features of a cytoplasmic membrane/cell wall anchorage motif common in over 20 other Gram-positive bacterial surface proteins. The features of the sorting signal are labeled.

TABLE I  
Aligned consensus sequences at the amino and carboxyl termini of proteins from *Streptococcus* and *Staphylococcus* species  
 $\beta$ -GlcNAc'ase is the abbreviation for  $\beta$ -N-acetylglucosaminidase. Residues in other *Streptococcal* and *Staphylococcal* proteins identical to  $\beta$ -N-acetylglucosaminidase are in boldface and conservative amino acid substitutions are underlined.

Source (Ref.)	Protein	N terminus	C terminus
<i>S. pneumoniae</i>	$\beta$ -GlcNAc'ase	K Q Q R F S I R K Y A V G A A S V L I G	L P E T G T
<i>S. agalactiae</i> (20)	$\alpha$ antigen	K Q R F S I K R F K F G A A S V L I G	L P A T G E
<i>S. agalactiae</i> (21)	$\beta$ antigen	K M R Y S I R K F S V G V A S V A V A	L P Y T G V
<i>S. pyogenes</i> (22)	<i>m</i> -related protein precursor	K K N Y S L R K L K T G T A S V A V A	L P S T G E
<i>S. pyogenes</i> (23)	IgA receptor precursor	K Q Y S L R K L K T G T A S V A V A	L P S T G E
<i>S. suis</i> (24)	Surface protein	K Q Q F S I R K Y H F G A A S V L L G	L P N T G E
<i>S. aureus</i> (25)	Fibronectin binding protein	N N L R Y G I R R H K L G A A S V F L G	L P E T G G
<i>S. aureus</i> (26, 27)	Protein A	K K N I Y S I R K L G V G I A S V T L G	L P E T G E
<i>S. epidermidis</i> (28)	Lipase	R Q N K Y S I R K F S V G A S S I L I A	L P P T I T

smaller length enzymes from clones pGXStrH17 and pGXStrH8, which were 2–2.5 times greater (Table III). The rates of hydrolysis, however, followed an opposite trend, where the  $V_{max}$  measured for the smallest enzyme clone pGXStrH8, was 4 times faster than the wild-type enzyme and 3 times faster than the largest recombinant enzyme from clone pGXStrH17.

The  $K_m$  and  $V_{max}$  was also determined for the factor Xa cleaved enzymes from clones pGXStrH8 and pGXStrH17 (Table III). The  $K_m$  of the fusion protein, from clone pGXStrH17, and the Xa-cleaved enzyme, StrH17, were very similar. By contrast, the  $K_m$  values of the clone pGXStrH8 and StrH8 recombinant

enzymes differed by a factor of two. However, the standard error of the  $K_m$  of the StrH8 enzyme was quite large (Table III).

**Aglycon Specificity of Recombinant  $\beta$ -N-Acetylglucosaminidase**—Affinity-purified fusion proteins were assayed for their ability to hydrolyze a panel of radiolabeled oligosaccharide alditol substrates to compare aglycon specificity with the wild-type purified  $\beta$ -N-acetylglucosaminidase (Fig. 5). The reaction products from enzyme digestions were separated by HPAEC. Because this technique shows a shift in retention times between the substrate and hydrolysis product indicating cleavage has occurred, the structures of hydrolysis products were veri-

TABLE II  
Hexosaminidase consensus sequence comparison between six different species

Boldface residues indicate perfect matching with the *S. pneumoniae*  $\beta$ -N-acetylglucosaminidase. Underlined residues indicate conservative amino acid substitutions relative to the *S. pneumoniae*  $\beta$ -N-acetylglucosaminidase.

Organism (Ref.)	Protein	Hexosaminidase consensus sequence
<i>S. pneumoniae</i>	$\beta$ -N-acetylglucosaminidase	<b>T Q A E V T E L I E Y A K S K D I G L I P A I N S P G H M D</b>
<i>S. pneumoniae</i>	StrH8'ase	<b>T E S Q M T D L I N Y A K D K G I G L I P T V N S P G H M D</b>
<i>Vibrio harveyi</i> (29)	N,N'-diacetylchitobiase	<b>S K A D Y V E I L K Y A K A R N I E V I P E I D M P A H A R</b>
<i>Vibrio vulnificus</i> (30)	$\beta$ -N-acetylhexosaminidase	<b>T R E D Y K E I L A Y A S A R N I Q V I P S M D M P G H S L</b>
<i>Homo sapiens</i> (31)	$\beta$ -hexosaminidase $\alpha$ -subunit	<b>T A Q D V K E V I E Y A R L R G I R V L A E F D T P G H T L</b>
<i>M. musculus</i> (32)	$\beta$ -N-acetylhexosaminidase	<b>T A Q D V K E V I E Y A R L R G I R V L A E F D T P G H T L</b>
<i>D. discoideum</i> (33)	$\beta$ -N-acetylhexosaminidase	<b>S H D D I Q E V V A Y A K T Y G I R V I P E F D I P G H A A</b>
<i>Candida albicans</i> (34)	$\beta$ -N-acetylhexosaminidase	<b>S K N D L K Y I V D Y A R A R G V R V T P E I D M P G H A R</b>

TABLE III  
Kinetic analysis of wild-type and recombinant *S. pneumoniae*  $\beta$ -N-acetylglucosaminidase

A desialylated, degalactosylated biantennary glycan was incubated with recombinant enzymes from clone pGXStr8 and -17, the corresponding factor Xa-cleaved enzyme, StrH8 and -17 or wild-type  $\beta$ -N-acetylglucosaminidase. Kinetic measurements were made as described in the text. The length of the pGX constructs includes 232 amino acids from the 26-kDa glutathione S-transferase fusion protein tag.

Enzyme	Amino acids (number)	Molecular mass <sup>a</sup>	$K_m$	$V_{max}$
		kDa	$\mu M$	$\mu mol/min/mg$
Wild-type $\beta$ -GlcNAc'ase	1311	120	$132 \pm 11$	$17 \pm 3.3$
pGXStrH17 $\beta$ -GlcNAc'ase	1067	121	$302 \pm 130$	$24 \pm 3.7$
pGXStrH8 $\beta$ -GlcNAc'ase	612	69	$364 \pm 88$	$75 \pm 6.9$
StrH17 $\beta$ -GlcNAc'ase (Xa-cleaved)	835	95	$278 \pm 89$	$113 \pm 13$
StrH8 $\beta$ -GlcNAc'ase (Xa-cleaved)	380	43	$723 \pm 510$	$197 \pm 73$

<sup>a</sup> The molecular mass of each enzyme was determined by SDS-PAGE.

fied using Bio-Gel P-4 analysis, which gave actual sizes of the reaction products in glucose units. The minimum enzyme concentration (units/ml), using 4-MU-GlcNAc for each enzyme that was chosen showed no hydrolysis of (GlcNAc $\beta$ 1-4GlcNAc)<sub>3</sub> and thus retained exclusive  $\beta$ 1-2 GlcNAc hydrolysis. Recombinant pGXStrH8 enzyme, from the most truncated clone, at 0.025 units/ml and wild-type  $\beta$ -N-acetylglucosaminidase at 0.012 units/ml, hydrolyzed the biantennary alditol to completion (Fig. 5a). Only the  $\beta$ 1-2-linked GlcNAc residues were removed from a triantennary glycan (Fig. 5b). However, only partial hydrolysis of a tetraantennary alditol was observed using the recombinant pGXStrH8 enzyme compared with wild-type enzyme (Fig. 5c), and no GlcNAcs were removed from either bisected-biantennary or bisected-hybrid oligosaccharide substrates (Figs. 5, d and e). Concentrations of 0.25 units/ml of recombinant pGXStrH8 enzyme were required to hydrolyze completely the susceptible  $\beta$ 1-2-linked GlcNAc of tetraantennary structures, but even at 45 units/ml, bisected oligosaccharides remained refractory to digestion. By contrast, the recombinant enzymes from the longest and intermediate clones pGXStrH17 and pGXStrH7, respectively, at low concentrations (0.025 units/ml) showed the same specificity as the wild-type enzyme, but only partially hydrolyzed those oligosaccharides with a bisecting GlcNAc. At a concentration of 1.5 units/ml, recombinant enzyme from pGXStrH17 exhibited the same activity as the wild-type enzyme when assayed using both bisected-biantennary and bisected-hybrid substrates. At higher enzyme concentrations (5 units/ml), this recombinant enzyme removed both terminal GlcNAc residues and the bisecting GlcNAc from bisected-biantennary substrates, an activity also achievable at high concentrations (0.1 units/ml) of the wild-type enzyme. The recombinant enzyme from clone pGXStrH7 exhibited partial activity against the bisected substrates at the highest enzyme concentration tested (17 units/ml). Thus the recombinant enzyme with the greatest number of amino acids, pGXStrH17 (Fig. 2, 1067 amino acids including the fusion tag, see Table III) shared the same aglycon specificity as the wild-type  $\beta$ -N-acetylglucosaminidase, whereas the smallest length clone, pGXStrH8 (612 amino acids, Table III) was completely unable to hydrolyze  $\beta$ 1-2-linked GlcNAc residues from bisected

oligosaccharide alditols at extremely high enzyme concentrations (the maximum tested was 120 units/ml of factor Xa-cleaved enzyme). The intermediate length clone pGXStrH7 (641 amino acids, see Fig. 2) was partially active in hydrolyzing bisected substrates. A summary of these results is shown in Fig. 6.

Recombinant enzymes that had the glutathione S-transferase fusion tag removed by cleavage with factor Xa were checked for activity against the five substrates used in Fig. 5 to confirm that the fusion tag did not affect the activity of the recombinant fusion proteins assayed above. At the same enzyme concentration, the factor Xa-cleaved recombinant enzymes showed identical aglycon specificities to the uncleaved enzymes (results not shown), demonstrating that the fusion tag did not change the substrate specificity of the recombinant enzymes.

#### DISCUSSION

The peculiar aglycon specificity exhibited by  $\beta$ -N-acetylglucosaminidase from *S. pneumoniae* not only makes it a useful tool for oligosaccharide sequencing, it also makes it an interesting candidate for the investigation of substrate-enzyme interactions. At low enzyme concentrations, this enzyme will only hydrolyze  $\beta$ 1-2-linked GlcNAc residues and is restricted by further N-acetylglucosaminidase substitutions of the  $\alpha$ 1-6 mannose arm of N-glycans and by bisecting GlcNAc residues of the core mannose. Little is known about the factors that might govern this restricted activity, though it might partly be explained by the potential steric hindrance created by the flexibility of 1-6-linked sugars, which have the additional  $\omega$  angle of rotation. However, because these restricted  $\beta$ 1-2-linkages are cleaved by higher concentrations of  $\beta$ -N-acetylglucosaminidase as well as by other hexosaminidases, the intrinsic properties of the enzyme itself, which correlate with this narrow substrate specificity, remain to be clearly defined. In this paper we have shown that *S. pneumoniae*  $\beta$ -N-acetylglucosaminidase is encoded by a unique 3933-bp gene, *strH*, which terminates in an ochre stop codon and possesses typical Gram-positive bacterial transcription initiation and termination sequences. The computer-translated amino acid sequence revealed two consensus motifs at both the amino and carboxyl terminus that are

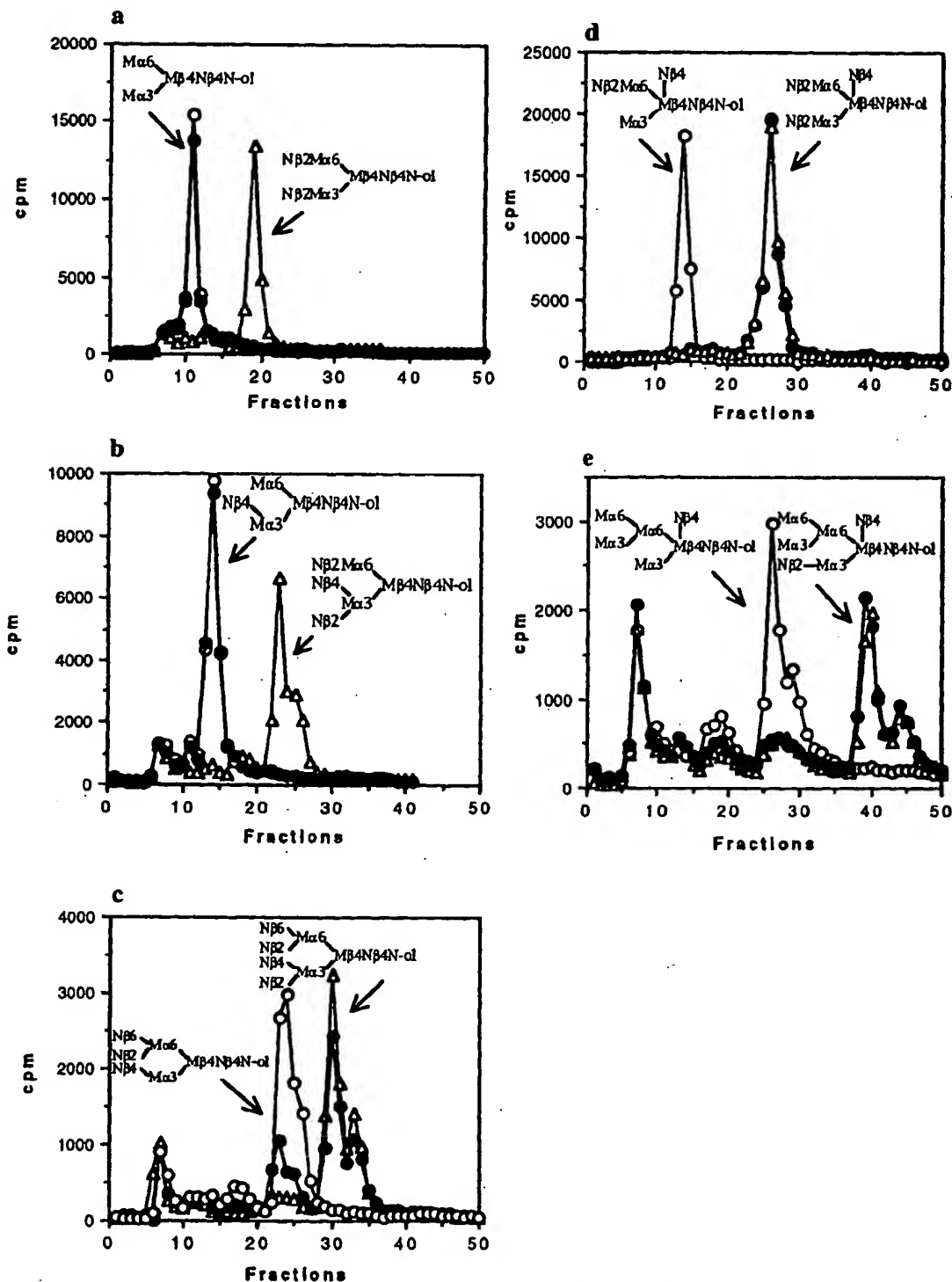


FIG. 5. Aglycon Specificity of  $\beta$ -N-acetylglucosaminidase from *S. pneumoniae*. Both substrate and hydrolysis product elution positions are shown for each oligosaccharide. N, GlcNAc; M, Man. The  $^3\text{H}$ -labeled oligosaccharide alditol substrates ( $2.5 \times 10^5$  cpm,  $\Delta$ ) were incubated with 12.5 milliunits/ml purified wild-type  $\beta$ -N-acetylglucosaminidase (O) and 25 milliunits/ml recombinant enzyme from clone pGXStrH8 (●) in 50 mM citric acid/sodium phosphate buffer, pH 5.0, containing 1 mg/ml bovine serum albumin at 37 °C for 18 h. Samples were desalted, and the hydrolysis products were separated by HPAEC using an eluant of 200 mM NaOH. Fractions (1 ml) were collected and scintillation counted for radioactivity.

common to other Gram-positive surface proteins. A tandem repeat was identified in the *strH* gene, and within each repeat lies a stretch of 30 amino acids homologous to sequences in six

other hexosaminidases found in a wide variety of species, suggesting that perhaps these amino acids may be important for the catalytic function of the enzyme. The tandem repeat re-

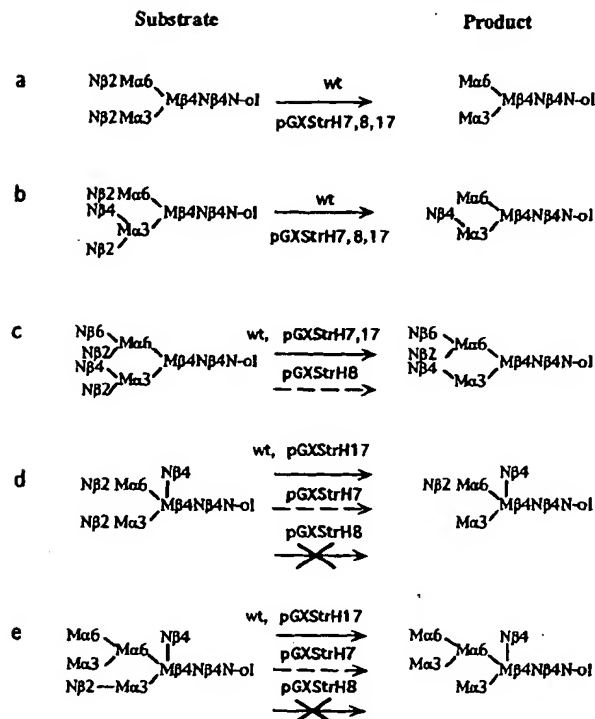


FIG. 6. Oligosaccharide structures hydrolyzed by *S. pneumoniae*  $\beta$ -N-acetylglucosaminidase. A summary of the data presented on the aglycon specificity of wild-type and recombinant  $\beta$ -N-acetylglucosaminidase. Solid arrows, complete hydrolysis at appropriate enzyme concentration; dashed arrows, partial hydrolysis; crossed arrows, no hydrolysis at all enzyme concentrations tested. See text and Fig. 5 for full details. N, GlcNAc; M, Man.

gions of protein may be correlated with the ability of the clones to hydrolyze substrate. Substrate specificity experiments revealed that the longest enzyme clone, pGXStrH17, was able to hydrolyze  $\beta$ 1-2-linked GlcNAc residues of bisected oligosaccharides similar to the wild-type enzyme. The intermediate sized cloned enzyme from pGXStrH7 demonstrated only a partial ability to hydrolyze these substrates, whereas the enzyme from shortest clone, pGXStrH8, was completely deficient in this activity. At the protein level, the differences between these clones are clearly defined by the presence or absence of the second tandem repeat region. Recombinant enzyme from pGXStrH17 contains both tandem repeats; pGXStrH7 contains all of the first and 95 amino acids (of 330) of the second tandem repeat region; and pGXStrH8 contains only the first tandem repeat. Sequence alignment revealed that these tandem repeat regions each contain a 30-amino acid consensus sequence present in six other hexosaminidases from divergent species, which leads us to speculate that this region may be part of an extended site of the enzyme, which is required for substrate orientation around the active site. In situations where this portion is partially missing, as with the shortest clone, efficient association with the some substrates may be affected and hydrolysis impaired. Further insights into the function of these tandem repeat regions may be gained by the separate expression of the second tandem repeat region followed by experiments addressing the substrate specificity. However, it is noteworthy that expression cloning failed to identify a clone that only contained the second repeat region and possibly indicates that the presence of this domain alone is not sufficient to maintain catalytic activity. The weaker affinity and increased rate of hydrolysis of biantennary oligosaccharides, together

with changes in aglycon specificity, correlate with an amino acid-specific extension of the polypeptide carboxyl terminus. This suggests that there is either a minimum amino acid residue requirement for active site conformation, or a requirement for protein domains outside the active site to recognize and direct substrates into the enzymes catalytic center.

A better understanding of the cellular location of the *Streptococcal*  $\beta$ -N-acetylglucosaminidase can be derived from the sequence information of the gene. The identification of a characteristic *Streptococcal* secretion signal at the amino terminus and a carboxyl-terminal sorting motif found in a variety of Gram-positive cell-surface proteins predicts a membrane-bound enzyme with an active site extending into the extracellular space. This structural model explains the origin of the  $\beta$ -N-acetylglucosaminidase purified from the media. Several lines of evidence support the hypothesis that this form of the enzyme originates from the cell surface and is released during protease-assisted autolysis. First, the cell-associated enzyme was found to be present in much greater quantities than the enzyme found in the medium over an 8-h growth period (3). Purification of the wild-type enzyme from *Streptococcal* cells revealed multiple amino termini, and the presence of a number of isoforms when examined by native-polyacrylamide gel electrophoresis (3). The major protein and activity stained band of this preparation analyzed by SDS-polyacrylamide gel electrophoresis migrates with a molecular mass of 120 kDa, 24 kDa smaller than the 144-kDa translated gene, providing further evidence that proteolysis, additional to signal peptide cleavage, occurs during the release of the cell-associated enzyme.

Second, only a single gene was obtained from expression cloning. If more than one  $\beta$ -N-acetylglucosaminidase was expressed by this organism, then more than one gene might have been isolated by enzyme activity. Although 32 clones were selected by their  $\beta$ -N-acetylglucosaminidase activity, further analysis of five of these selected at random revealed that they were overlapping fragments of the same gene.

This evidence, together with the substrate specificity data (3), implies with great certainty that both forms of the  $\beta$ -N-acetylglucosaminidase are derived from the same protein located at the cell-surface.

The contribution to substrate configuration made by amino acid residues outside the active site in glycohydrolytic enzymes is a largely unexplored phenomenon. The availability of truncated enzymes that have altered substrate specificities as described here provides a unique opportunity to study the mechanism by which glycosidases bind and hydrolyze complex oligosaccharides.

**Acknowledgments**—We thank Professor Raymond Dwek for support. We also thank Ned Siegel and Christine Smith, Monsanto Co. St. Louis, for providing amino-terminal amino acid analysis.

#### REFERENCES

- Hughes, R. C., and Jeanloz, R. W. (1964) *Biochemistry* 3, 1543-1548
- Glasgow, L. R., Paulson, J. C., and Hill, R. L. (1977) *J. Biol. Chem.* 252, 8615-8623
- Clarke, V. A., Willenbrock, F. W., Scudder, P., Rotasert, J., Jacob, G. S., and Butters, T. D. (1994) *Int. J. Biochem.* 1, 151-168
- Krivan, H. C., Roberts, D. D., and Ginsburg, V. (1988) *Proc. Natl. Acad. Sci. U. S. A.* 85, 6167-6161
- Linder, T. E., Daniels, R. L., Lim, D. J., and Demaria, T. F. (1994) *Microb. Pathog.* 16, 435-441
- Boulnois, G. J. (1992) *J. Gen. Microbiol.* 138, 249-259
- Yamashita, K., Ohkura, T., Yoshima, H., and Kobata, A. (1981) *Biochem. Biophys. Res. Commun.* 100, 226-232
- Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Southern, E. (1976) *J. Mol. Biol.* 98, 503-517
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463-5467
- Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* 162, 156-159
- Smith, D. B., and Johnson, K. S. (1988) *Gene (Amst.)* 87, 31-40
- Butters, T. D., Scudder, P., Rotasert, J., Petrusson, S., Fleet, G. W. J.,

- Willenbrock, F. W., and Jacob, G. S. (1991) *Biochem. J.* **279**, 189-195
14. Robbins, P. W., Overbye, K., Albright, C., Benfield, B., and Pero, J. (1992) *Gene (Amst.)* **111**, 69-76
15. Goward, C. R., Scawen, M. D., Murphy, J. P., and Atkinson, T. (1993) *Trends Biochem. Sci.* **18**, 136-140
16. Abrahamsen, L., Moks, T., Nilsson, B., Hellman, U., and Uhlen, M. (1985) *EMBO J.* **4**, 3901-3908
17. von Heijne, G. (1984) *J. Mol. Biol.* **173**, 243-251
18. Schneswind, O., Model, P., and Fischetti, V. A. (1992) *Cell* **70**, 267-281
19. Braun, V., and Sieglin, U. (1970) *Eur. J. Biochem.* **13**, 336-348
20. Michel, J. L., Madoff, L. C., Olson, K., Kling, D. E., Kasper, D. L., and Ausubel, F. M. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 10060-10064
21. Jerlstroem, P. G., Chhatwal, G. S., and Timmis, K. N. (1991) *Mol. Microbiol.* **5**, 843-849
22. Haanes, E. J., and Cleary, P. P. (1989) *J. Bacteriol.* **171**, 6397-6408
23. Friths, E., Heden, L. O., and Lindahl, G. (1989) *Mol. Microbiol.* **3**, 1111-1119
24. Smith, H. E., Vecht, U., Gielkens, A. L., and Smits, M. A. (1992) *Infect. Immun.* **60**, 2361-2367
25. Signäs, C., Ravcci, G., Jönsson, K., Lindgren, P. E., Anantharamaiah, G. M., Höök, M., and Lindberg, M. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86**, 699-703
26. Loeftdahl, S., Guss, B., Uhlen, M., and Lindberg, M. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80**, 697-701
27. Uhlen, M., Guss, B., Nilsson, B., Gatenbeck, S., Philipson, L., and Lindberg, M. (1984) *J. Biol. Chem.* **259**, 1695-1702
28. Farrell, A. M., Foster, T. J., and Holland, K. T. (1993) *J. Gen. Microbiol.* **139**, 267-277
29. Soto-Gil, R. W., and Zyakind, J. W. (1989) *J. Biol. Chem.* **264**, 14778-14783
30. Somerville, C. C., and Colwell, R. R. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6751-6755
31. Myerwitz, R., Piekarz, R., Neufeld, E. F., Shows, T. B., Jr., and Suzuki, K. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 7830-7834
32. Beccari, T., Hoade, J., Orlacchio, A., and Stirling, J. L. (1992) *Biochem. J.* **285**, 593-596
33. Graham, T. R., Zassenhaus, H. P., and Kaplan, A. (1988) *J. Biol. Chem.* **263**, 16823-16829
34. Cannon, R. D., Niimi, K., Jenkinson, H. F., and Shepherd, M. G. (1994) *J. Bacteriol.* **176**, 2640-2647